网络隐私争议事件中用户隐私关注及情感对比研究*

■ 谭芳[1] 杨阳[1] 卓伊玲[1] 徐健[1] 肖卓[2]
[1] 中山大学资讯管理学院 广州 510006 [2] 中山大学图书馆 广州 510275

摘 要：[目的/意义]人工智能、大数据等领域的快速发展使得商业发展与隐私保护之间的矛盾愈发尖锐。通过对不同类型的网络隐私争议事件微博评论进行情感及话题对比分析，以探究不同情境下网络用户的隐私态度的异同点与背后机理。[方法/过程]采集2012年至2019年网络隐私争议事件的相关微博评论，对其进行预处理，作为实验数据；基于情感词典计算各评论的情感强度值，并将隐私争议事件分为隐私收集类、隐私曝光类及隐私协议类，对比分析不同情境下的用户评论情感趋势；构建用户隐私讨论对象-情感表达二分网络，并通过二分网络投影构建单顶点网络，结合节点中心性等指标进行二分网络及投影分析。[结果/结论]结果表明，用户整体隐私关注呈现上升趋势；不同类型隐私争议事件的用户负面情感强度水平不同；不同隐私争议情境下用户的关注热点差异较大，情感表达各有特点。以上结果表明不同情境中的用户隐私关注及情感表现具有明显差异。

关键词：网络隐私 隐私关注 情感分析 二分网络
分类号：G203
DOI：10.13266/j.issn.0252-3116.2021.02.009

大数据、人工智能等新兴领域的快速发展使得基于用户数据的关联隐私信息挖掘成为可能，这些因素的共同作用为隐私安全带来重大隐患。国家互联网应急中心发布的《2019年上半年我国互联网网络安全态势》指出"移动APP强制授权、过度索权、超范围收集个人信息的现象大量存在，违法违规使用个人信息的问题十分突出"[1]。在隐私泄露现象如此严峻的情形下，部分用户采取提供虚假信息甚至拒绝提供个人信息的方式以保护个人隐私[2-3]。然而，对于企业而言，用户的隐私披露对科技发展、业务开展和商业价值的实现都具有重要意义。因此，对用户隐私态度及行为等方面的研究便显得十分重要。

目前，信息领域的隐私研究主要聚焦于某一群体在社交媒体、大数据、定向广告中的隐私关注和行为的影响因素或机理探究[4-6]。有研究表明，隐私认知和行为高度依赖于情境[7]，即在不同用户特征、文化背景、商业环境、平台类型等不同情境下，用户的隐私认知和行为有所不同[8-10]。本研究通过对网络隐私争议事件进行整理和分类，基于用户评论进行隐私关注及情感特征对比分析，以探究不同情境下互联网用户的隐私态度的异同点，为企业及相关决策者的隐私实践工作提供理论参考。

1 文献回顾

"隐私"最早从法律角度被定义为"免受外界干扰的独处权利"[11]。目前为止，已形成从权利说、商品说、控制说和状态说四种角度对隐私的定义[12-14]。隐私的控制说源自信息控制理论，如A. F. Westin提出的"个人、团体或社团自主决定由于自身信息在何时，以何种方式，在何种程度上传播给他人的权利"[15]。控制说在网络隐私定义和隐私关注研究中应用较为广泛[16]。然而，隐私内涵始终随着信息技术的发展而不断变化，需要采用一种更包容的定义方式[13]。沟通隐私管理理论（Communication Privacy Management）由S. Petronio提出[17]，其从控制论的角度对隐私进行定义，并认为隐私的边界是动态的，在网络隐私研究中得到广泛应用[18-19]。基于该理论，本文将网络隐私定义为"个人认为属于自己的可进行传播范围控制和管理的

* 本文系广东省自然科学基金项目"情感分歧度量化模型及其应用研究"（项目编号：2018A030313981）研究成果之一。
作者简介：谭芳（ORCID：0000-0003-3007-0547），硕士研究生；杨阳（ORCID：0000-0003-0154-1637），硕士研究生；卓伊玲（ORCID：0000-0001-7522-6328），硕士研究生；徐健（ORCID：0000-0003-4886-4708），教授，硕士生导师；肖卓（ORCID：0000-0003-4531-4292），馆员，硕士，通讯作者，E-mail：xiaozhuo@mail.sysu.edu.cn。
收稿日期：2020-06-25 修回日期：2020-09-07 本文起止页码：87-97 本文责任编辑：杜杏叶

在网络中呈现的信息”。

隐私关注是指“人们对隐私信息可能被泄露的主观认知”[20]。在信息学领域,对隐私关注的研究包括影响因素与行为意向两方面。李贺等基于解释水平理论,发现隐私关注仅对社交媒体用户的远期隐私披露意图产生影响[21]。沈旺等则发现社交媒体用户的隐私关注会负向影响其近期和远期披露意愿[22]。但二者研究的情境都仅限于社交媒体,其结论在其他情境中是否适用,仍然值得探究。陈小燕等对 LBS 情境下的用户隐私关注作用机制进行了研究,发现在移动服务环境中,隐私关注对披露意愿的负作用仍然成立[23]。因此,有学者指出,当前较有意义的研究已经从用户的披露意愿研究转变为了隐私入侵情境下用户的不同反应[24]。一些学者对隐私关注中的个体差异进行了研究,如 E. Van den Broeck 等将成年阶段分为三个阶段,发现中年成年组比新兴成年与年轻成年组更关注自身隐私[25]。杨嫚等研究了用户对精准广告的隐私关注现状,发现不同年龄、学历的用户呈现不同的隐私关注水平,“19 岁以下和专科及以下学历的用户的隐私关注水平偏低”[26]。然而,这些研究往往以某一平台的使用情况为例,并没有对不同情境中的隐私关注进行探究和区分。高山川等区分了网络平台类型,发现网络平台的不同类型对用户授权意愿影响不同[10]。L. Yu 等考虑了平台类型对用户隐私披露的影响,发现用户在工具型平台中感知隐私风险对披露意愿的负面影响远大于情感型平台[7]。这类研究证实了不同情境下的用户隐私认知及行为存在差异,但除了平台类型,其他情境如在隐私侵入方式和受侵犯的隐私信息类型等不同的情况下,用户隐私关注及认知表现差异仍然处于未知状态,这些差异能够为互联网行业的隐私实践提供一定参考。

隐私争议事件作为加剧用户隐私关注的潜在事件,是进行隐私关注研究非常有价值的情境和数据来源,在不同类型事件发生后,用户在情感、认知层面的反应十分值得探究[27]。基于此,本研究以 2012 年至 2019 年期间发生的网络隐私争议事件及其微博评论为研究对象,通过对现有文献进行梳理,根据不同隐私侵入方式对事件进行分类,在研究中试图回答以下几个问题:①不同网络隐私争议情境中,用户的隐私关注变化趋势如何?②不同网络隐私争议情境中,用户是否呈现不同的隐私情感特征?③不同网络隐私争议情境中,用户的隐私关注热点是否存在差异?存在哪些差异?本文通过考察和对比不同情境下用户的情感特征与关注热点,并结合二分网络与单顶点网络分析方法,挖掘背后的差异机理,其创新之处在于:对不同情境下用户的隐私关注和隐私态度的情感维度进行分析,为基于大众媒体的隐私分析提供了新的视角。

2 研究设计

本文研究设计可分为六部分:数据获取及处理、情感值计算、网络节点提取、情感值对比、二分网络构建和用户隐私关注热点及情感特征对比,如图 1 所示:

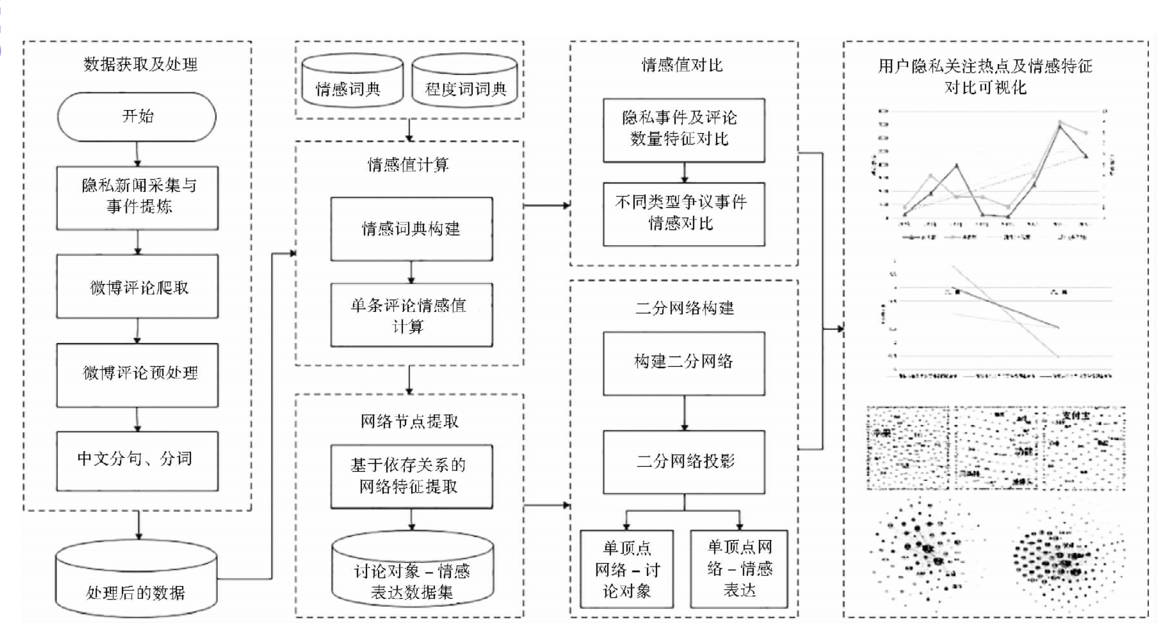


图 1 网络隐私争议事件中用户隐私关注及情感对比研究设计

本文重点考察能给用户带来成本与收益矛盾的事件，即隐私争议事件，并将“隐私争议”定义为“引起用户隐私披露，并在其风险与收益权衡之间产生矛盾而引起的争议”。因此，所关注的事件主要为处于法律监管空白地带的隐私争议事件，以下研究步骤基于该定义进行。

2.1 数据获取与处理

将中国新闻网[28]作为主要新闻事件采集来源，搜狐新闻平台[29]和新浪微博平台[30]作为补充，以关键词“应用 隐私”“APP 隐私”“社交 隐私”“科技 隐私”进行检索。采集标准为：①以事件性新闻为主；②剔除网络诈骗、黑客攻击等违法事件。采集时间为 2019 年 10 月。以采集到的事件为单位，在新浪微博平台进行事件检索，其评论需满足：①有效数量大于 30 条；②以隐私讨论为主，且讨论立场呈现多样化。最终得到 2012 年至 2019 年共 31 个事件。其次，使用八爪鱼采集器[31]爬取微博评论，并进行预处理，包括去重、评论去噪等。最终得到评论数 20 471 条，演变情况如图 2 所示：

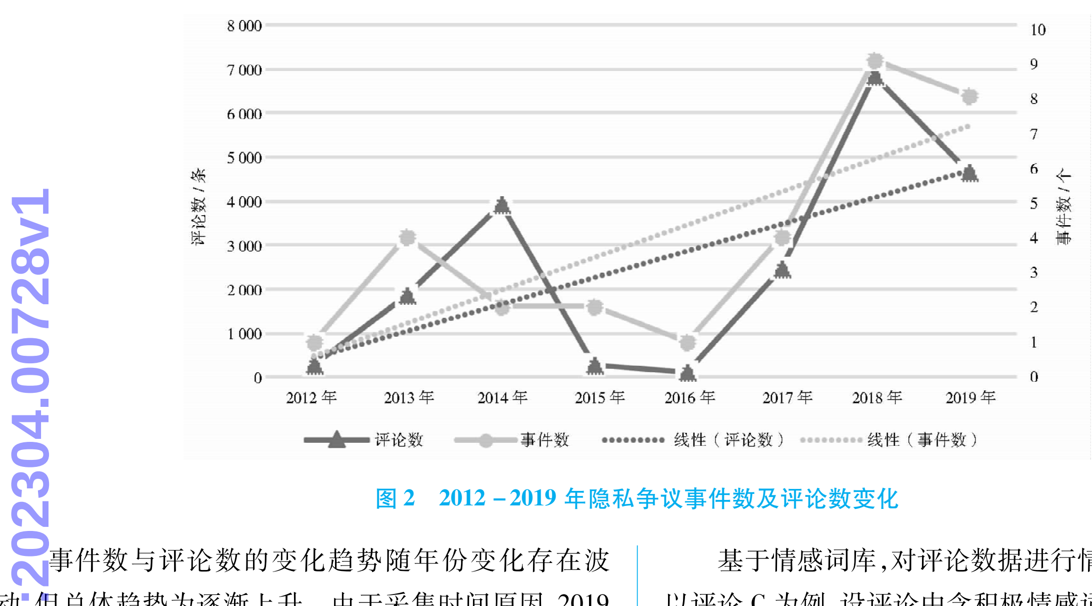


图 2 2012－2019 年隐私争议事件数及评论数变化

事件数与评论数的变化趋势随年份变化存在波动，但总体趋势为逐渐上升。由于采集时间原因，2019 年数据收集不完整，较 2018 年略有下降，属正常现象。各年份中的隐私争议事件和评论数量存在不同幅度的波动，如在 2015 年和 2016 年中，筛选出的隐私争议事件较少，而曝光的网络诈骗、信息买卖等网络安全违法事件较多，可能的解释是与信息安全政策的实施、网络安全工作的大力开展和网络实名制的普及等因素相关。因此，为了使实验结果更具代表性，考虑到评论数量的波动情况和时间间隔的均衡性，将 2012 年到 2015 年归为一个时期，称为“第Ⅰ时期”，其事件数为 9 个，共 6 360 条评论；将 2016 年到 2019 年归为一个时期，称为“第Ⅱ时期”其事件数为 22 个，共 14 111 条评论。

2.2 情感值计算与情感值对比

以大连理工大学的情感词汇本体库作为基本情感词典[32]，百度文库的弹幕情感词典作为补充，去除其中能表达情绪但无法表达用户对隐私看法的情感词[33]，并根据评论数据添加了表达隐私看法的情感词。此外，收集常用程度副词和否定词形成相应词典，最终得到情感词[34]、常用程度副词和否定词[35]以共同构成本研究的情感词库。

基于情感词库，对评论数据进行情感强度值计算。以评论 C 为例，设评论中含积极情感词 n 个，消极情感词 m 个，情感强度值计算如公式(1)所示：

$$Senti_C = \sum_{i=0}^{n}(D_i T_i + E_i) - \sum_{j=0}^{m}(D_j T_j + E_j) \quad \text{公式(1)}$$

其中，$Senti_C$ 为评论 C 的情感强度值；i 和 j 分别为评论中所涉及的积极情感词序列数和消极情感词序列数；D 对应序列数情感词前的副词或否定词的程度得分，若无则为 1；T 对应序列数情感词的情感得分；E 对应情感词后的感叹号得分，若无感叹号则为 0，有则为 2，以增强权重[36]。

根据现有文献对隐私争议事件进行分类，统计其情感强度值变化，并进行对比分析，可反映用户在不同情境下对待隐私的情感特征，有助于不同领域的隐私实践决策，其演化趋势也可为用户隐私态度趋势预测提供参考。

2.3 网络节点提取及二分网络构建

二分网络是复杂网络中的一种重要存在形式，由两类节点及其连接构成，仅允许不同组别的节点相连[37]。将二分网络进行投影，可形成由各类别节点组成的两组单顶点网络，以反映同类节点之间的关系，其

示意图见图 3[38]。二分网络可揭示更深层的网络特性，在个性化推荐[39]、作者合著关系研究[40]等多个领域被应用。本文通过二分网络对各类型隐私争议事件进行可视化呈现，构建讨论对象 – 情感表达节点关系，以了解不同类型隐私争议事件中用户的隐私关注和情感呈现规律。

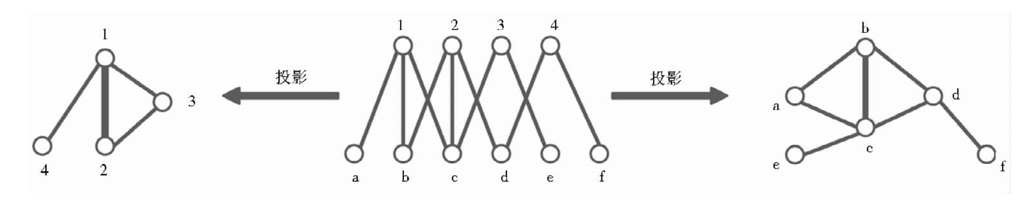


图 3 二分网络及其节点投影示例

本文实施其步骤为：

（1）对用户评论进行高频词统计，筛选出隐私“讨论对象”与“情感表达”词汇，形成相应词典[41]，基于该词典，使用哈工大 LTP 工具抽取出讨论对象 – 情感表达词对，构成二分网络中的两组节点，最终抽取出 1 206组关系形成关联词表，作为原始数据集。

（2）基于数据集，使用 Gephi 复杂网络分析软件[42]进行二分网络构建，并选用合适参数和聚类方法以完善可视化效果。

（3）为了进一步探究用户隐私关注中，“讨论对象”和“情感表达”的核心呈现规律，将数据集导入社会网络分析工具 Pajek 软件[43]，将二分网络以加权投影的方式转化为单顶点网络，并使用 Gephi 工具[42]进行可视化。

通过对各类型隐私争议事件中用户的关注主体和情感表达的对比分析，进一步挖掘用户在不同情境中隐私关注的具体内容与情感表达的异同点，分析背后差异机理，为企业相关决策者提供理论参考。

3 用户隐私关注及情感特征对比

3.1 隐私事件分类

国内外学者尝试从不同角度对网络隐私进行分类，早期较典型的是 J. K. Burgoon 将隐私分为四个维度，并认为网络隐私包含信息隐私、社会隐私和心理隐私，他总结到当时相关研究主要集中于对信息隐私的研究，信息隐私包括医药记录、财政记录、应用程序数据等[44]。然而，网络隐私事件中，企业对隐私的“侵犯”往往涉及多种类型，因此根据隐私类型对网络隐私事件进行维度划分并不合适。

2016 年出台的《中华人民共和国网络安全法》中，第四章对网络信息安全制度做了全面规定[45]。根据本文第 1 部分对网络隐私的定义，可以发现条款中第四十条 – 第四十五条与网络隐私侵犯事件直接相关。其中，第四十一、四十三条对网络运营者收集且使用用户信息进行了规定；第四十二条对网络运营者将用户信息提供给第三方做了规定；第四十四条对第三方窃取、倒卖个人信息行为做了规定[45]。这几条法律主要针对隐私信息的收集、使用、窃取和倒卖。结合整理的事件，将网络隐私争议事件分为以下三种类型，如表 1 所示：

表 1 隐私争议事件类型划分

事件类型	具体释义	举例
“隐私收集类”争议事件	指组织/产品/服务/某项科技涉嫌未经用户允许便收集用户个人数据，或将收集的信息用于精准营销、科技研发或提供给第三方等各种商业目的，从而引起隐私争议的事件。	“cookies 跟踪用户行为为痕迹引隐私争议” “qq 聊天窗口输入”荒野行动“弹出吃鸡游戏广告，qq 被质疑监测用户聊天记录”
“隐私曝光类”争议事件	指组织/产品/服务/某项科技未经用户允许便公开用户个人数据，以达到各种商业目的，从而引起隐私争议的事件。	“吃外卖点了啥，好友都知道！外卖应用绑定社交被指侵犯隐私” “360 摄像头网上直播引发隐私争议 吃饭健身可一览无余”
“隐私协议类”争议事件	指组织/产品/服务/某项科技因隐私协议不规范，或是强制或默认授权而直接引起的隐私争议事件。	“‘ZAO’app 用户隐私协议不规范争议” “支付宝默认勾选芝麻信用协议”

以下内容将采用表 1 所示的隐私争议事件类型划分标准进行分析。

3.2 隐私事件及评论数量特征对比

根据表 1 划分的类型对 2.1 小节中获取的 31 个事件进行分类，各类型事件数及评论数变化趋势见图 4。

由图 4 可以看出，各类事件数量均随时期上升，而评论量的变化各有特点。“隐私收集类”争议事件的数量在两个时期均为最多，其次为“隐私曝光类”，最后为“隐私协议类”。然而，“隐私收集类”争议事件的

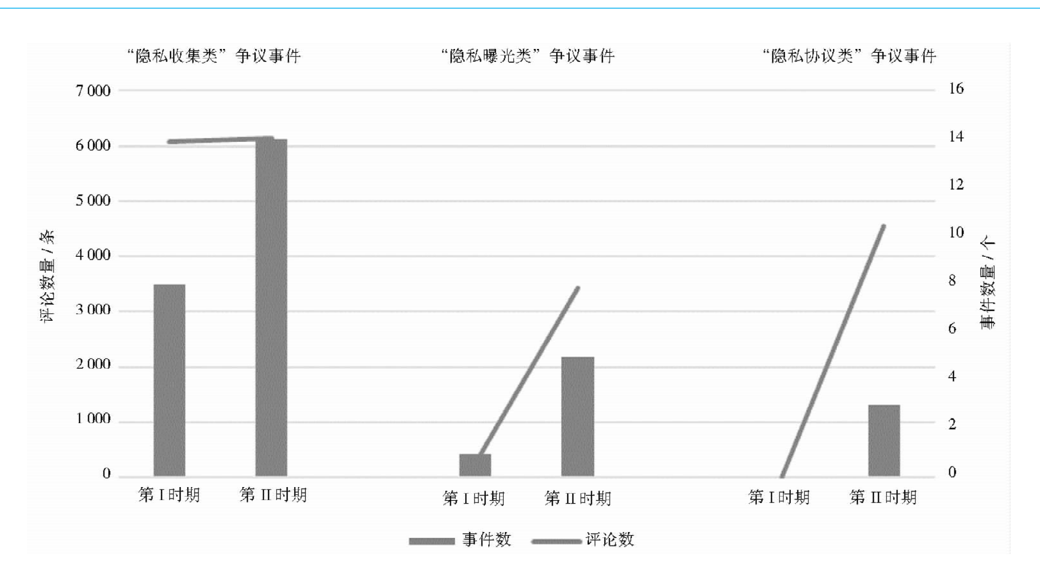


图 4　各时期不同类型隐私争议事件数量及评论数量变化

事件数量虽有较大增长,但评论数量增长并不明显,主要是由于在第Ⅰ时期出现了一例产生较大影响的隐私事件“苹果首次承认可提取短信与照片”,评论数占据64%;“隐私协议类”争议事件在第Ⅱ时期才出现,且评论数量直接赶超“隐私曝光类”,是隐私争议事件中的新型热点事件。

3.3　不同类型争议事件情感值对比

对用户在不同隐私争议事件中的情感演变进行对比,变化趋势如图5所示:

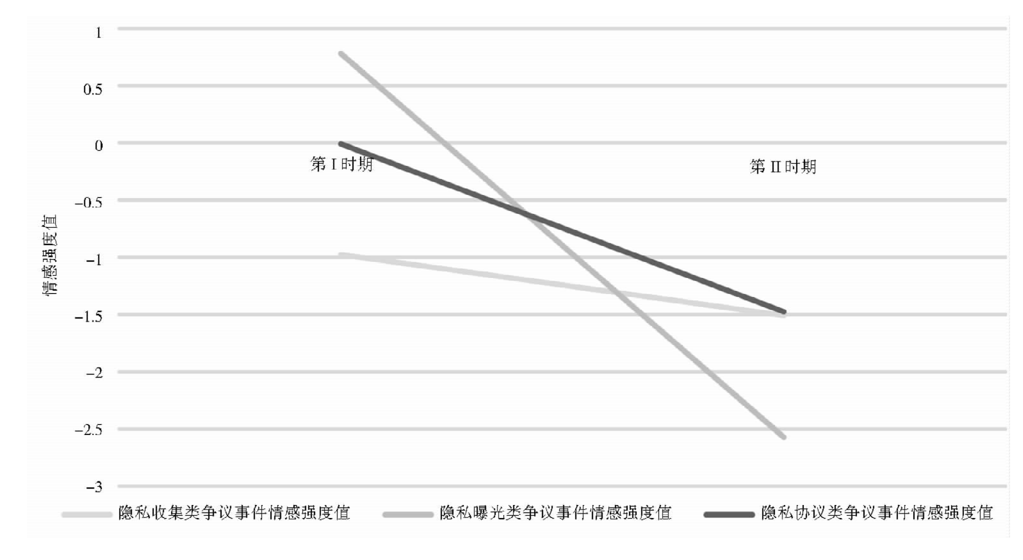


图 5　各时期不同类型隐私争议事件情感强度值演变趋势

各类型事件早期评论情感均呈下降趋势,其中,“隐私曝光类”事件的情感强度值从第Ⅰ时期的0.79变为第Ⅱ时期的-2.57,“隐私收集类”事件情感强度值从-0.96降到-1.50,“隐私协议类”争议事件在早期并未出现,是新型隐私争议事件,其情感强度值从0下降到-1.47,与“隐私收集类”事件基本达到一致。

负面情感的增强往往意味着用户隐私关注的上升[20]。据此可知,不同类型事件的用户隐私关注均呈上升趋势,这与大数据应用的泛滥和隐私侵犯事件的频繁曝光相关,即媒体负面曝光的数量变化对用户隐私关注程度造成正面影响,与郭龙飞的结论一致[46]。其中,尽管“隐私收集类”事件数最多,但其隐私关注幅度变化却是最小,而隐私关注上升幅度最大的为“隐私曝光类”事件,这说明,不同情境中的用户在隐私关注上所表现的差异与媒体负面曝光的静态数量占比没有显著相关性,可能的解释是由于用户隐私收集现象长期以来一直存在,且在各阶段的发生频次最高,用户需要为隐私保护所付出的努力成本变高,有效性降低,导致用户“隐私疲劳”程度更深,而“隐私疲劳”能够促进隐私消极行为的发生[47],如对

滥用个人信息持“无所谓”的态度等，如评论“无所谓，本来就是透明的”。

3.4　不同类型隐私事件关注热点及情感对比

该部分通过构建二分网络及各节点投影，以进一步了解不同类型隐私争议事件用户的隐私关注热点和情感呈现规律。

3.4.1　二分网络分析

根据 2.3 所述步骤，构建三种类型争议事件的讨论对象－情感表达二分网络，见图 6。图 6 中，节点大小按“度”排序，节点越大说明被评论次数越多；边的大小与颜色深浅按“边的权重”显示，边越粗，说明两节点被共同提及的次数越多。

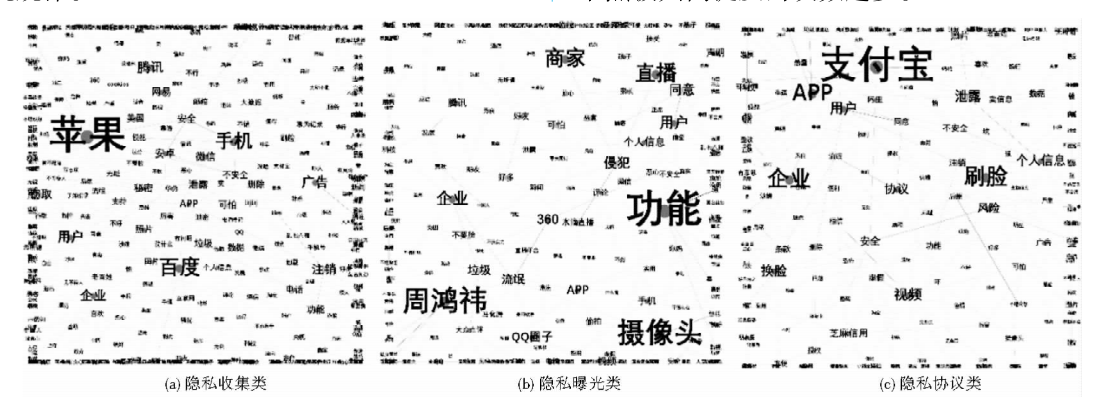


图 6　各类型争议事件讨论对象－情感表达二分网络图

“隐私收集类”争议事件中，节点较大的为“苹果”“手机”等，情感表达中节点较大的为“注销”“窃取”等，结合评论原文，发现用户对 APP 注销问题十分关注，如评论“你注销了也不代表你的信息被删除了，还是存在别人的数据库中的”，表明用户对被遗忘权的关注。“隐私曝光类”争议事件中，节点较大的为“功能”“摄像头”等，情感表达中，主要节点包括“垃圾”“流氓”等，反映出用户的厌恶情感，相关评论如“的确被侵犯了隐私，出门还被直播，＊＊就是垃圾”等。“隐私协议类”争议事件中，节点较大的为“支付宝”“刷脸”等，情感表达中，较大的节点如“泄露”“安全”等，该情境中用户尤其关注生物信息的隐私安全问题，相关评论如“支付宝刷脸我都不敢用”。

三类事件中，用户的隐私关注热点差异较大。在情感表达上，用户对“隐私曝光类”争议事件的情感表达较“隐私收集类”与“隐私协议类”而言更为负面，与 3.3 中情感对比结论一致。解释水平理论（construal level theory）认为，当某项活动不属于个人的直接经历时，在心理上便是遥不可及的[48]。“隐私收集类”与“隐私协议类”争议事件并未造成即时且可感知的危害后果，用户对这类情境下侵犯隐私的风险是假设性的，属于远心理距离。而“隐私曝光类”事件中，用户对隐私侵犯风险的感知是清晰的，为近心理距离。陈小燕等发现在移动服务环境下，用户的感知披露风险将增加其隐私关注水平[23]。由此可知，不同情境中，心理距离越近，即对侵犯隐私的感知风险越清晰时，用户产生的负面情感越强，隐私关注越明显。这对陈小燕等人的结论起到了补充作用[23]。

3.4.2　讨论对象－情感表达单顶点网络分析

二分网络无法直观反映隐私涉及的讨论对象与情感表达之间的直接联系，因此，抽取出“讨论对象”和“情感表达”节点，通过加权投影得到单顶点网络，可进一步反映用户对两组节点的关注分布情况。图 7 至图 9 为三类隐私事件的讨论对象和情感表达单顶点网络。节点间的权重为共同连接节点的数量，节点的不同颜色灰度代表不同聚类类别。为了优化展示效果，对节点进行适当合并，仅选取度值为 1 及以上的节点。

（1）“隐私收集类”事件。中介中心性反映了行动者在中间路径的控制能力[49]。对图 7（a）“讨论对象”聚类后，发现“手机”的中介中心性最高（535.66），是连接各社团的核心词。由图密度可看出，各社团联系较为密切。通过适当合并可分为三类社团，即①以“手机”“苹果”等为核心，节点占比 58%，涉及对通讯设备和工具中隐私收集的争议，如评论“苹果只能苹果公司窃取，安卓机所有 app 都可以窃取，我还是选择苹果”；②以“百度”“微信”为核心，节点占比 21%，涉及搜索引擎和社交软件隐私收集的讨论，如评论“微信也会读取你的聊天记录来精准推送广告”；③以“数据”“APP”为核心，节点占比 18%，讨论对象涉及电商和生活服务平台，关注的隐私信息类型包括“手机号”“银行卡”等，如评论“手机号码一登陆什么信息都泄漏了”。

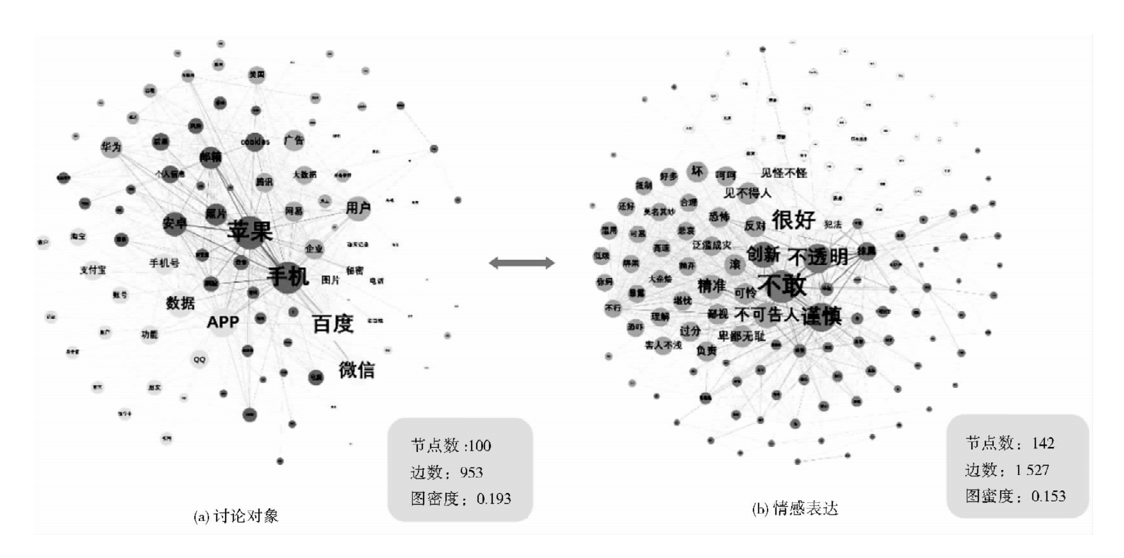


图7　“隐私收集类”事件讨论对象－情感表达单顶点网络

图7(b)“情感表达”单顶点网络中,“不敢”的中介中心性达1 122.77,是连接各社团的核心词。该网络同样可分为三类,即:①以“不敢”“谨慎”为核心的社团,节点占比47.18%,以恐惧类情感为主;②以“不可告人”“卑鄙无耻”等为核心的社团,节点占比27.46%,以厌恶情感为主;③以“很好”“见怪不怪”为核心的社团,节点占比25.35%,以正面或中性情感为主。

综上,“隐私收集类”事件中用户的主要关注点为通讯设备的隐私收集问题,涉及的讨论对象包括设备本身与设备上的各类应用程序,以恐惧类情感为主,并带有厌恶和积极两类矛盾的情感。借助“社会交换理论”[50]的分析视角,用户通过让渡浏览、互动和购物等信息,能够换取到更具个性化和便捷化的服务。但随着该类型的营销策略被科技企业广泛运用,用户对隐私风险的感知愈加强烈,在这种情况下,由于用户无法直接监督企业的信息使用情况,导致隐私风险往往带有较大的不确定性,因此,越来越多用户对自己隐私保护行为的有效性产生了质疑,即隐私保护自我效能降低,导致恐惧类情感占比上升,而其他类型情感占比相应下降。

(2)“隐私曝光类”事件。如图8(a)所示,“直播”的中介中心性最高(272.74),是连接“讨论对象”聚类后各社团的核心词。对社团进行适当合并,可分为两大类:①以“直播”“摄像头”“个人信息”等为核心,节点占比55.32%,回顾评论原文,发现用户的关注点在于公众场所的监控被作为直播,及其与个人隐私泄露之间的争议,如评论“摄像头安装于公共场所,无所谓开不开直播”“我可不愿意吃个饭还被人直播”;②以“企业”“周鸿祎”“大众点评”等为核心,节点占比21.27%,主要指各类生活服务等应用的隐私曝光行为,如评论“美团还可以看到微信好友点了什么外卖吓得都不敢点了”。

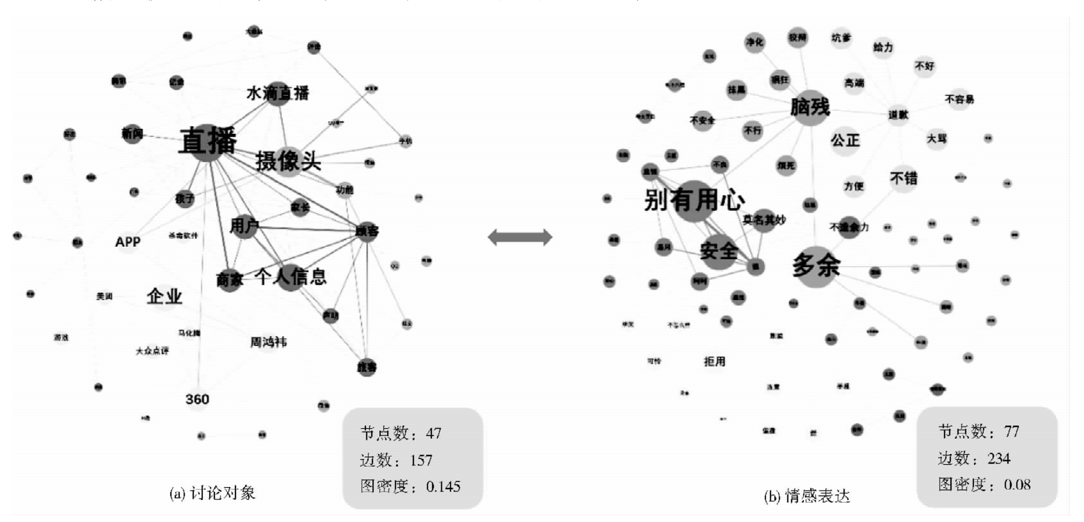


图8　“隐私曝光类”事件讨论对象－情感表达单顶点网络

在图 8(b)“情感表达”单顶点网络中,“多余”是连接各社团的核心词,其中介中心性达 449.13。各社团通过合并可分为:①以“多余”“脑残”“拒用”为核心,节点占比 37.66%,结合评论数据,是对各讨论对象功能的评价,说明对用户而言其功能价值不高,不值得以隐私为代价;②以“别有用心”“安全”为核心,节点占比 22.08%,体现的是用户对隐私安全的谨慎态度;③以“不错”“道歉”为核心,节点占比 12.99%,是平台或企业对公众道歉后,用户的原谅态度,如评论“认错的态度倒是不错”。综上,“隐私曝光类”事件的主要关注热点在于“直播”等社交娱乐类应用,用户所表现的负面情绪较为强烈。一方面,与隐私收集类事件不同的是,在该类事件中,用户对隐私泄露和使用情况的感知是清晰的,例如在监控直播事件中“泄露”的各种监控视频,从而导致了用户强烈的不满和质疑情绪。另一方面,由于用户隐私“共享”的受众群体是其他用户而非受法律和协议制约的企业,所造成的隐私不可控的风险更大。此外,基于隐私计算理论,用户在面临隐私问题时,会对预期收益和所感知的隐私风险进行评估,从而做出隐私行为[51]。在该情况下,用户会下意识对产品价值(主要为产品功能)进行衡量,当认为所获收益低于出售隐私成本,但却被迫承担隐私风险时,则会产生强烈的厌恶和抵抗情感。

(3)“隐私协议类”事件。如图 9(a)所示,对“讨论对象”进行聚类,“支付宝”是连接各社团的核心词,其中介中心性达 334.92。将社团进行适当合并,分为两类:①以“支付宝”“用户”“风险”等为核心词,节点占比 75%,结合评论原文,体现的是用户对用户协议不规范所造成隐私风险的关注,如评论“支付宝这样的协议明显是存在严重的个人隐私安全风险”。②以“个人信息”“刷脸”为核心词,节点占比 17.31%,涉及对生物信息隐私的关注,如评论“一张脸部照片能泄露的生物特征是很多的,至少市面的部分安卓手机能解开,所以慎用吧”。

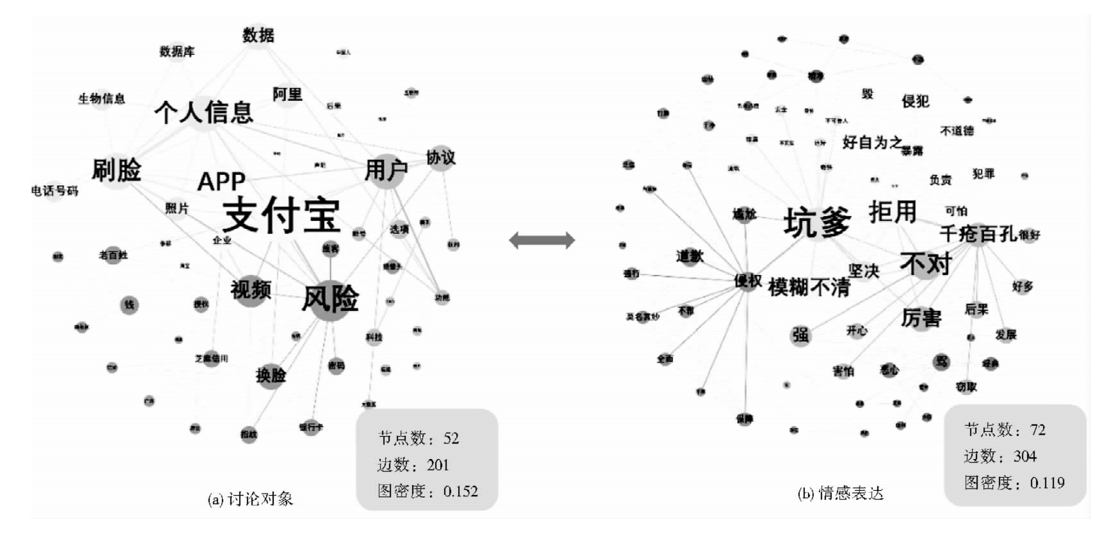


图 9 “隐私协议类”事件讨论对象 – 情感表达单顶点网络

在图 9(b)“情感表达”单顶点网络中,“坑爹”为连接各社团的核心词,其中介中心性达 822.59。对社团适当合并,可分为:①以“坑爹”“侵权”“拒用”为核心,节点占比 53.42%,体现用户的厌恶及抵抗情感;②以“千疮百孔”“不对”“厉害”为核心,节点占比 27.40%,回顾评论原文,反映了用户对企业侵犯隐私做法的谴责和对产品价值的认可,体现出不同用户态度的争议性。

综上,“隐私协议类”事件中,用户对生活服务类应用的隐私问题较为关注,以负面情绪为主。L. Yu 将平台类型分为情感型平台和工具型平台,前者以社交及娱乐等这类存在情感关系互动的平台为主,而后者主要以实现交易的平台为主[7]。本研究中,隐私收集类和隐私曝光类事件中的涉及的讨论主体以情感型应用为主,而隐私协议类事件中主要涉及平台以工具型应用为主。与情感平台相比,工具型平台中用户对于各项隐私的披露往往具有特定目的,因此会表现得更加理性,对平台的隐私政策的审查和成本效益的评估也会更加谨慎[7]。用户协议作为用户隐私的基本书面保障,在不规范的情况下,会触发多数用户的厌恶情感和抵抗行为,该现象也从另一视角验证了 E. S. Wang 的结论,即用户对隐私政策的感知有效性降低则会显著降低用户使用行为的信心,并直接对其行为产生影响[52]。

4 结论与不足

互联网络的迅猛发展使得网络隐私问题愈加严重。现有研究主要围绕隐私关注成因及行为意向展开,对不同情境中的用户隐私关注及情感研究较少。本文重点关注处于法律监管空白地带的隐私争议事件,以 2012 年至 2019 年网络隐私争议事件及微博评论为研究对象,对不同情境下的用户隐私关注、情感及关注热点进行差异分析,包括评论数据获取与处理、情感值计算、网络节点提取、情感值对比、二分网络构建和用户隐私关注热点及情感特征对比。本研究对隐私关注领域的研究视角起到一定补充作用,其结论可为互联网等行业的相关决策者提供实践参考。主要结论如下:

(1)不同类型隐私争议事件的用户隐私关注均呈上升趋势。不同类型事件之间,用户在隐私关注上所表现的差异与媒体负面曝光的静态数量占比没有显著相关性。基于现有文献,将隐私争议事件分为“隐私收集类”争议事件、“隐私曝光类”争议事件和“隐私协议类”争议事件。其中,“隐私收集类”争议事件数量(22 件)及评论数量(12 223 条)在各时期均占比最大,情感强度值降幅却最小,下降了 56%,隐私关注变化幅度相对不大,可归因于该情境中,用户需要为隐私保护所付出的努力成本变高和有效性降低所导致的“隐私疲劳”现象;“隐私曝光类”争议事件情感强度值降幅最大,下降超过 400%,情感强度值最低。

(2)不同类型隐私争议事件的用户负面情感强度水平不同。基于解释水平理论,当心理距离越近,即对侵犯隐私的感知风险越清晰时,用户在隐私争议事件中的负面情感越强,隐私关注越强。“隐私曝光类”事件中,用户对隐私泄露情况的感知是清晰的,而在“隐私收集类”与“隐私协议类”事件中,用户的隐私风险感知往往基于假设性,因此,前者表现的负面情感较后两者更强烈。

(3)用户在不同情境下的隐私关注热点差异较大,情感表达各有特点。在“隐私收集类”争议事件中,用户首要关注通讯设备的信息收集问题,但由于用户无法直接监督企业的信息使用情况,所感知的隐私风险往往带有较大的不确定性,因此以恐惧类情感为主;“隐私曝光类”争议事件中,关注点为社交娱乐类应用的隐私泄露问题,由于所感知的隐私泄露情况更为清晰,隐私曝光的受众群体更不可控,并结合产品价值与隐私成本的衡量等因素,用户表现出强烈的厌恶类情感;“隐私协议类”争议事件中,涉及的讨论主体以工具型平台为主,用户表现更为理性和谨慎,其对隐私政策感知有效性的降低对用户情绪产生了显著负面影响,同样以厌恶类情感为主,从另一视角验证了 E. S. Wang 的结论[52]。

本文的不足之处有如下几点:首先,微博评论仅代表一部分用户的隐私态度,并未考虑其他平台的用户和非网民群体,因此实验结论存在一定局限性;其次,本文只是将所有网络用户看成一个整体,今后可从不同群体网络隐私情感及关注热点的表现差异等方面开展进一步研究。

参考文献:

[1] 国家计算机网络应急技术处理协调中心. 2019 年上半年我国互联网网络安全态势[R/OL]. [2020 -03 -24]. http://www.cac.gov.cn/2019-08/13/c_1124871484.htm.

[2] 王乐,王璐瑶,孙早. 隐私侵犯经历对网络用户自我披露的影响机制[J]. 系统工程理论与实践,2020,40(1):79 -92.

[3] MARTIN K D, MURPHY P E. The role of data privacy in marketing[J]. Journal of the academy of marketing science,2017,2(45):135 -155.

[4] 于婷婷,杨蕴焓. 精准广告中的隐私关注及其影响因素研究[J]. 新闻大学,2019(9):101 -116.

[5] 刘茜. 青年用户微信朋友圈隐私管理动因研究[J]. 当代传播,2019(4):84 -89.

[6] ALASHOOR T, HAN S, JOSEPH R C. Familiarity with big data, privacy concerns, and self-disclosure accuracy in social networking websites: an APCO model[J]. Communications of the association for information systems,2017,41(4):62 -96.

[7] YU L, LI H, HE W, et al. A meta-analysis to explore privacy cognition and information disclosure of internet users[J]. International journal of information management,2020,51: 1 -10.

[8] JEONG Y, COYLE E. What are you worrying about on Facebook and Twitter? An empirical investigation of young social network site users' privacy perceptions and behaviors[J]. Journal of interactive advertising,2014,2(14):51 -59.

[9] JEONG Y, KIM Y. Privacy concerns on social networking sites: interplay among posting types, content, and audiences[J]. Computers in human behavior,2017,69:302 -310.

[10] 高山川,王心怡. 网络平台和收益的类型对信息隐私决策的影响[J]. 应用心理学,2019,25(4):364 -371.

[11] WARREN S D, BRANDEIS L D. The right to privacy[J]. Harvard law review,1890,5(4):193 -220.

[12] 段伟文,纪长霖. 网络与大数据时代的隐私权[J]. 科学与社会,2014,4(2):90-100.
[13] 刘健,陈卓."隐私"研究的三种话语分析[J]. 陕西理工大学学报(社会科学版),2017,35(2):47-51.
[14] 吕耀怀,熊节春. 信息隐私界定的 BIT 模式[J]. 图书馆理论与实践,2011(6):35-39.
[15] WESTIN A F. Privacy and freedom[M]. New York: Athenum, 1967.
[16] 李睿. 移动互联网环境下隐私泄露容忍度的测量与实证研究[D]. 大连:大连理工大学,2014.
[17] PETRONIO S. Boundaries of privacy: dialectics of disclosure[M]. Albany, New York: State University of New York Press, 2002.
[18] 钟瑛,刘利芳. 信息传播中的隐私侵犯及保护[J]. 新闻与写作,2018(2):23-26.
[19] EINSPÄNNER J. Privacy online: perspectives on privacy and self-disclosure in the social Web[J]. New media & society, 2013, 8(15):1401-1402.
[20] 陈昊,李文立,柯育龙. 社交媒体持续使用研究:以情感响应为中介[J]. 管理评论,2016,28(9):61-71.
[21] 李贺,余璐,许一明,等. 解释水平理论视角下的社交网络隐私悖论研究[J]. 情报学报,2018,37(1):1-13.
[22] 沈旺,高雪倩,代旺,等. 基于解释水平理论与调节定向理论的社交网络隐私悖论研究[J]. 情报科学,2020,38(8):120-127,140.
[23] 陈小燕, CLIQUET G. 隐私权衡理论视角下用户隐私关注研究[J]. 技术经济与管理研究,2020(5):9-13.
[24] MARTIN K D, MURPHY P E. The role of data privacy in marketing[J]. Journal of the academy of marketing science, 2017, 45(2):135-155.
[25] VAN DEN BROECK E, POELS K, WALRAVE M. Older and wiser? Facebook use, privacy concern, and privacy protection in the life stages of emerging, young, and middle adulthood[J]. Social media + society,2015,1(2):671657067.
[26] 杨嫚,温秀妍. 隐私保护意愿的中介效应:隐私关注、隐私保护自我效能感与精准广告回避[J]. 新闻界,2020(7):41-52.
[27] 谢毅,高充彦,童泽林. 消费者隐私关注研究述评与展望[J]. 外国经济与管理,2020,42(6):111-125.
[28] 中国新闻网[EB/OL]. [2020-03-24]. https://www.chinanews.com/.
[29] 搜狐新闻[EB/OL]. [2020-03-24]. https://news.sogou.com.
[30] 新浪微博[EB/OL]. [2020-03-24]. https://weibo.com.
[31] 八爪鱼采集器[EB/OL]. [2020-03-24]. https://www.bazhuayu.com/.
[32] 徐琳宏,林鸿飞,潘宇,等. 情感词汇本体的构造[J]. 情报学报,2008,27(2):180-185.
[33] 邓淑卿,李玩伟,徐健. 基于句法依赖规则和词性特征的情感词识别研究[J]. 情报理论与实践,2018,41(5):137-142.
[34] 情感词典[EB/OL]. [2020-03-24]. https://figshare.com/articles/12240224.
[35] 常用程度副词及否定词词典[EB/OL]. [2020-03-24]. https://figshare.com/articles/12240233.
[36] 徐健,吴思洋. 网络用户评论的情感分歧度量化算法研究[J]. 情报学报,2020,39(4):427-435.
[37] 卢伟聪,徐健. 基于二分网络的网络用户评论情感分析[J]. 情报理论与实践,2018,41(2):121-126.
[38] LATAPY M, MAGNIEN C, VECCHIO N D. Basic notions for the analysis of large two-mode networks[J]. Social networks,2008,30(1):31-48.
[39] 李树青,徐侠,许敏佳. 基于读者借阅二分网络的图书可推荐质量测度方法及个性化图书推荐服务[J]. 中国图书馆学报,2013,39(3):83-95.
[40] 张金柱,韩涛,王小梅. 作者-关键词二分网络中的合著关系预测研究[J]. 图书情报工作,2016,60(21):74-80.
[41] 隐私讨论对象-情感表达词典[EB/OL]. [2020-03-24]. https://figshare.com/articles/___-___/12401426.
[42] BASTIAN M, HEYMANN S, JACOMY M. Gephi:an open source software for exploring and manipulating networks[C]//The Association for the Advancement of Artificial Intelligence. Proceedings of the third international conference on weblogs and social media. Menlo Park,California:The AAAI Press,2009:361-362.
[43] BATAGELJ V, MRVAR A. Pajek-program for large network analysis[J]. Connections,1998,2(21):47-57.
[44] BURGOON J K, PARROTT R, LE POIRE B A, et al. Maintaining and restoring privacy through communication in different types of relationships[J]. Journal of social and personal relationships, 1989,2(6):131-158.
[45] 全国人民代表大会常务委员会. 中华人民共和国网络安全法[EB/OL]. [2020-03-24]. http://www.moj.gov.cn/Department/content/2016-11/23/592_201322.html.
[46] 郭龙飞. 社交网络用户隐私关注动态影响因素及行为规律研究[D]. 北京:北京邮电大学,2013.
[47] CHOI H, PARK J, JUNG Y. The role of privacy fatigue in online privacy behavior[J]. Computers in human behavior, 2018, 81:42-51.
[48] TROPE Y, LIBERMAN N, WAKSLAK C. Construal levels and psychological distance:effects on representation,prediction,evaluation,and behavior[J]. Journal of consumer psychology,2007,2(17):83-95.
[49] BRANDES U. A faster algorithm for betweenness centrality[J]. Journal of mathematical sociology,2001,2(25):163-177.
[50] 李秉权. 社交媒体语境下隐私侵权、边界划定及保护研究[D]. 长春:吉林大学,2020.
[51] 刘英华,翟嘉靖. 信息系统用户隐私计算研究综述[J]. 大学图

书情报学刊,2020,38(3):113 -119.
[52] WANG E S. Effects of brand awareness and social norms on user-perceived cyber privacy risk[J]. International journal of electronic commerce,2019,23(2):272 -293.

作者贡献说明：
谭芳:设计研究方案,采集数据,进行实验,论文撰写;
杨阳:数据预处理;
卓伊玲:采集数据,情感词典完善;
徐健:设计研究方案,论文最终版修订;
肖卓:提出研究思路,论文最终版修订。

Comparison of Privacy Concern and Sentimental Characteristics of Users in Internet Privacy Controversial Events

Tan Fang[1] Yang Yang[1] Zhuo Yiling[1] Xu Jian[1] Xiao Zhuo[2]
[1] School of Information Management, Sun Yat-Sen University, Guangzhou 510006
[2] Sun Yat-Sen University Library, Guangzhou 510275

Abstract: [**Purpose/significance**] The rapid development of AI, big data etc. makes an increasingly fierce confrontation between business development and privacy protection. This paper finds the differences in privacy concern and sentimental characteristics of users in different situations through the comparative analysis of sentiment and topic of Weibo comments on different types of online privacy controversial events. [**Method/process**] Firstly, Weibo comments related to privacy controversial events from 2012 to 2019 were collected and preprocessed as experimental data. Secondly, the sentiment value of each comment was calculated based on the sentiment dictionary. Then, the privacy controversial events were classified into privacy collection categories, privacy exposure categories and privacy agreement categories. The sentimental analysis was conducted under different situations. Finally, a bipartite network of "discussed object - emotional expression" was constructed, and a single vertex network was constructed by the bipartite network projection, and the analysis of both were carried out by combining the indexes such as node centrality. [**Result/conclusion**] The result shows that users' overall privacy concern shows an upward trend; users of different types of privacy controversial events have different levels of negative sentimental intensity; the hotspots of users' concern vary widely across privacy controversial events, and emotional expression have different characteristics. These findings indicate that there are significant differences in user privacy concern and sentimental performance in different situations.

Keywords: internet privacy privacy concern sentiment analysis bipartite network